

## Tema 3 — Actividad no. 2

### Cuestionario

31 de agosto, 2022

1. 2 aplicaciones que tiene el análisis léxico en el PLN

Dado que el léxico es el conjunto de palabras que constituyen una lengua, su análisis en el PLN, se centra en el reconocimiento y la categorización de estas palabras, para su posterior análisis. Mediante el análisis léxico es posible generar predictores de texto en la escritura, o bien correcciones gramaticales en un texto ya escrito.

2. ¿Para qué sirve la tokenización?

La tokenización es el conteo de las palabras que hay en un texto (entendiendo palabra como aquello que está entre dos espacios en blanco), así como el conteo de cada palabra diferente en el mismo texto. Sirve para obtener análisis estadísticos sobre la forma de distribución de las palabras en el texto, o las palabras más usadas por ejemplo, y con ello generar análisis lingüísticos.

3. Explica la diferencia entre un token y un type.

- **Token:** Es el conteo de cada palabra que aparece en el texto.
- **Type:** Es el conteo de cada palabra diferente que hay en el texto.

4. ¿En qué consiste el conteo de palabras?

Consiste en la separación individual de cada palabra, esta segmentación se hace a partir de la definición de palabra que se usa en el curso, posteriormente y según sea la forma de estudiar el texto, se lleva a una lista de palabras para calcular su frecuencia, es decir el número de veces que se repite cada una de las palabras en el corpus analizado. Según sea el caso se puede hacer distinción o no entre las palabras formadas con mayúsculas, y la misma palabra formada con minúsculas. En esta parte podemos contabilizar la frecuencia absoluta o relativa.

5. ¿Cómo podemos saber el número de veces que se repite una palabra en el corpus?

Para hacer esto, hacemos uso primero de una segmentación en una lista de palabras simple; es decir una lista de los types y su frecuencia absoluta, es decir que cada palabra con la cantidad de veces que se repite en el corpus.

6. ¿Qué es y qué papel juegan los lexicones computacionales en el PLN?

Los lexicones suelen ser las bases de trabajo para PLN. Ayudan para el análisis léxico, como sintáctico y semántico, así como a la desambiguación de los significados; mediante las relaciones entre palabras que se hacen en estos grandes diccionarios.

7. ¿Para qué sirven los N-gramas en PLN?

Sirven para generar otro tipo de información cuantificable sobre las palabras y las oraciones que se generan. Los N-Gramas se pueden convertir en representaciones estadísticas sobre las estructuras de palabras, sus secuencias, y la estructura de palabras.

8. ¿Qué son las colocaciones?

Las colocaciones son combinaciones de dos o más palabras que ocurren con más frecuencia de lo normal. Estas combinaciones suelen tener significados particulares, y se dan bajo ciertos contextos de la lengua, lo que hace que en particular en estas conjunciones no tenga sentido analizarlas de forma separada.

9. ¿En qué consiste la cuantificación de las colocaciones?

Para Cuantificar las colocaciones se toma en cuenta: El conteo de Palabras cercanas, El conteo de las palabras que aparecen hacia la izquierda, así como hacia la derecha, y el conteo de concordancias. Hacer la cuantificación de colocaciones ayuda a entender el contexto donde se usa, bajo que estructuras ligísticas se usan, y su significado en conjunto.

